Claims

1. A method for creating an assembly comprising:

(a) obtaining a set of DNA or RNA sequence reads;

(b) grouping the DNA or RNA sequence reads into categories;

(c) running an assembly program on each of the separate categories of DNA or RNA sequence reads; and

(d) repeating steps (b) and (c) as necessary.

2. The method of claim 1 wherein the grouping of the DNA or RNA sequence reads is done based on a function.

3. The method of claim 2 where said function is based on the size of the reads.

4. The method of claim 2 where said function is based on the entropy of the reads.

5. The method of claim 2 where said function is based on the GC percentage of the reads.

6. The method of claim 2 where said function is based on the nature of the longest repeat of the reads.

7. The method of claim 2 where said function is based on the nature of the one or more regions of highest entropy of the reads.

8. The method of claim 2 where said function is based on the nature of the one or more regions of lowest entropy of the reads.

9. The method of claim 2 where said function is based on the compression ratio of the reads.

10. The method of claim 2 where said function is based on the compression ratio of appended sequences of the reads.

11. The method of claim 2 where said function is a hybrid of two or more functions based on the qualities of the reads.

12. An apparatus for creating assemblies from DNA or RNA sequence reads comprising:

(a) means for obtaining a set of DNA or RNA sequence reads;

(b) means for grouping DNA or RNA sequence reads into categories; and

(c) means for creating assemblies from the DNA or RNA sequence reads.

13. In a computer system having a graphical interface including a display device and a selection device, a method of displaying information on the display device in a menu form and accepting menu selection input from a user, the method comprising:

retrieving a set of menu entries for the menu, each of the menu entries representing a method to perform upon DNA or RNA sequence reads;

displaying the set of menu entries on the display device;

displaying a set of parameters on the display device;

providing the user an opportunity to modify said set of parameters;

receiving an indication of a menu entry selection from the user via the selection device; and

in response to said indication of a menu entry selection, performing a method on the DNA or RNA sequence reads to create assemblies based on said set of parameters and said set of menu entries.

14. A set of application program interfaces embodied on a computer-readable medium for execution on a computer in conjunction with an application program that determines assemblies of DNA or RNA sequence reads, comprising:

a first interface that receives functions for a method for assembling DNA or RNA sequence reads;

a second interface that receives parameters for said functions;

a third interface that receives DNA or RNA sequence reads; and

returns an assembly of said DNA or RNA sequence reads.

15. A method for assembling sequence reads, comprising the steps of:

a) categorizing a plurality of sequence reads into at least two sub-groups of sequence reads based on an identifiable characteristic of the sequence reads in each sub-group;

b) matching sequences reads within each sub-group thereby creating assemblies of said sequence reads within each respective sub-group; and

c) repeating steps a) and b) with all unassembled sequence reads and newly created assemblies.

16. A method as set forth in claim 15, wherein said categorizing step includes identifying sequence reads having similar sizes.

17. A method as set forth in claim 15, wherein said categorizing step includes identifying sequence reads having similar entropies.

18. A method as set forth in claim 15, wherein said categorizing step includes identifying sequence reads having similar GC percentages.

19. A method as set forth in claim 15, wherein said categorizing step includes identifying sequence reads having similar longest repeats.

20. A method as set forth in claim 15, wherein said categorizing step includes identifying sequence reads having similar natures of regions of high entropy.

21. A method as set forth in claim 15, wherein said categorizing step includes identifying sequence reads having similar natures of regions of low entropy.

22. A method as set forth in claim 15, wherein said categorizing step includes identifying sequence reads having similar compression ratios.

23. A method as set forth in claim 15, wherein said categorizing step includes identifying sequence reads having compression ratios after sequence appending.

24. A method as set forth in claim 15, wherein said categorizing step includes identifying sequence reads having two or more similar characteristics.

25. A method as set forth in claim 15, wherein said categorizing step includes identifying sequence reads having